DOCUMENT RESUME

ED 340 742 TM 017 807

AUTHOR Trevisan, Michael S.; Sax, Gilbert
TITLE Estimating the Optimum Choice Format Using an Incremental Option Paradigm.
PUB DATE 91
NOTE 8p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Testing; Distractors (Tests); *Estimation (Mathematics); Grade 12; High Schools; High School Seniors; *High School Students; *Models; *Multiple Choice Tests; *Test Construction; Test Format; Test Items; Test Reliability
IDENTIFIERS *Internal Consistency

ABSTRACT

The purpose of this study was to compare the reliabilities of two-, three-, four-, and five-choice tests using an incremental option paradigm. Test forms were created incrementally, a method approximating actual test construction procedures. Participants were 154 12th-grade students from the Portland (Oregon) area. A 45-item test with two options per item was developed; and three-, four-, and five-option test forms were constructed by adding options to the two-option per item test. Reliability coefficients were calculated from the different forms of the test before and after implementing the Tversky condition, which assumes that testing time is proportional to test length. Despite significant differences in the reliability coefficients before invoking the Tversky condition, the magnitudes of the internal consistency reliability estimates for the three-, four-, and five-option formats after implementing the Tversky condition were similar, suggesting that time and energy might have been saved by constructing three-option items without loss in reliability. One table presents data about the test forms, and an 11-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340742

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MICHAEL S. TREVISAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Estimating the Optimum Choice Format Using an Incremental Option Paradigm

by

Michael S. Trevisan
RMC Research Corporation

Gilbert Sax
University of Washington

TM017807

BEST COPY AVAILABLE

The efficacy of the 3-option item format when constructing multiple-choice examinations has been demonstrated theoretically (e.g., Lord, 1944, 1977) and empirically (e.g., Costin, 1972; Straton & Catts, 1980; Trevisan, Sax, & Michael, In press). The methodological starting point for these studies is to first obtain a test constructed of items with the greatest number of options. Then, an option-elimination technique is employed (Haladyna & Downing, 1989a) to create forms of the test with fewer options per item. These forms are administered, and KR-20s were computed and compared.

Constructing different test forms used in these studies is both convenient and practical. However, it is clear that an interaction exists between the option-elimination technique chosen and the optimum number of options per item (Budesco & Nevu, 1985; Haladyna & Downing 1989a). It is difficult to account for this interaction (Budesco & Nevu, 1985) and it is an unavoidable by-product when using this method to construct different test forms.

Perhaps more troubling than the influence of the option-elimination technique on the reliability results of these studies is that constructing tests with different choice formats in this way is not the method used by test constructers. In actual test development, tests are built by adding alternatives to items until the desired number of options per item is reached (usually four or five). This methodological distinction is important and calls into question the validity of the results from option elimination studies. Perhaps more valid results would be

obtained if test forms with different choice formats were constructed using a method similar to those used in actual test construction.

Purpose

The purpose of the present study was to compare the reliabilities of 2-, 3-, 4- and 5-choice tests using an incremental option paradigm. It improves upon the design of previous studies that employ option-elimination techniques by creating test forms incrementally—a method that more closely approximates actual test construction procedures.

Method

One hundred and fifty-four twelfth grade students from the tri-county area of Portland, Oregon, participated in this study. A 45 item test with 2-options per item was developed using item-writing rules proposed by Roid and Haladyna (1982). The content of this test covered music, art, civics, geography, and history. Three-, 4-, and 5-option test forms were constructed by developing and adding options to the 2-option per item tests. The additional options were systematically constructed using the taxonomy of item-writing rules outlined by Haladyna and Downing (1989b), that include guidelines for distractor development. The four forms of the examination were randomly assigned to individual students, each student taking only one form of the test.

KR-20s were calculated for each test form and statistically compared using the M statistic (Hakstian & Whalen, 1976). This statistic is distributed as chi-

square and tests the null hypothesis that a group of KR-20s is obtained from the same distribution of reliability estimates. The Tversky condition (Tversky, 1964), which assumes that testing time is proportional to test length, has been incorporated in some studies (e.g., Green, Sax, & Michael, 1982; Lord, 1977; Trevisan, Sax, & Michael, In press). This assumption is controversial and has been shown not to hold in some settings (Budescu & Nevo, 1985). For this study, results after implementing the Tversky condition were calculated and comparisons made with results before implementing this condition. The Tversky condition is implemented by using the Spearman-Brown formula to estimate the reliability of a lengthened version of the 2-, 3-, and 4-option test forms given the total number of options found in the 5-option form. The KR-20s, means, standard deviations, sample sizes, and standard errors of measurement before and after implementing the Tversky condition can be found in Table 1.

Results

The results of the study showed nonsignificant differences existing among the reliability coefficients for the different forms of the test after implementing the Tversky condition ($\chi^2(3, \underline{N} = 154) = 2.76, p > 0.15$). The optimum number of options was four. The 0.15 significance level was invoked to reduce the probability of a type II error.

Discussion

The mean test scores range from a low of 18.55 for the 5-option version to

a high of 28.63 for the 2-option version, suggesting a fairly difficult test for these examinees. This in part may explain the low magnitude of the KR-20s (see table).

A large difference in the magnitude of the reliability coefficients was found between the 2-option test format with a KR-20 of 0.42 and the next highest KR-20 at 0.65 for the 3-option test format. The magnitude of this difference may in part account for the statistical significance found among the reliability coefficients before adjusting for the Tversky condition.

However, despite significant differences among the reliability coefficients before invoking the Tversky condition, the magnitude of the internal consistency reliability estimates for the 3-, 4-, and 5-option formats after implementing the Tversky condition were similar at 0.76, 0.79, and 0.71, respectively. Thus, considerable time and energy might have been saved by constructing 3-option items rather than 4- or 5-option items without loss in reliability. Also, if the assumption is made that testing time is proportional to test length (the Tversky condition) a large increase in the magnitude of the reliability coefficient for the 3-option test was found (from 0.65 before adjusting to 0.76 after adjusting). Another benefit of using the 3-option item when assuming proportionality is an increase in content validity over the 4- or 5-option item because more items can be created that measure content. Also, this study provides additional evidence for the efficacy of the 3-option item. By including more items having three options, increased reliability can be expected.

References

Budescu, D. and Nevo, B. (1985). Optimal number of options: an investigation of the assumption of proportionality. *Journal of Educational Measurement*, 22, 183-196.

Costin, F. (1972). Three-choice versus four-choice items: Implications for reliability and validity of objective achievement tests. *Educational and Psychological Measurement*, 29, 1035-1038.

Green, K., Sax, G., & Michael, W. (1982). Validity and reliability of tests having differing numbers of options for students of differing levels of ability. *Educational and Psychological Measurement*, 42, 239-245.

Hakstian, A.R. & Whalen, T.E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika*, 41, 219-231.

Haladyna, T.M. & Downing, S.M. (1989a). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2, 51-78.

Haladyna, T.M., & Downing, S.M. (1989b). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2, 37-50.

Lord, F.M. (1944). Reliability of multiple-choice tests as a function of choices per item. *Journal of Educational Psychology*, 35, 175-180.

Lord, F.M. (1977). Optimal number of choices per item-A comparison of four approaches. *Journal of Educational Measurement*, 14, 33-38.

Roid, G.H. & Haladyna, T.M. (1982). *A technology of test-item writing.* New York: Academic Press.

Straton, R.G. & Catts, R.M. (1980). A comparison of two-, three-, and four-choice item tests given in a fixed total number of choices. *Educational and Psychological Measurement*, 40, 357-365.

Trevisan, M.S. & Sax, G. (In press). The effects of the number of options per item and student ability on test validity and reliability. *Educational and Psychological Measurement.*

Table 1

Means, standard deviations, KR-20s, SEMs, and sample sizes for each test form before and after adjusting for the Tversky condition.

Form	No. of Persons	n of items	$\overline{X}$	SD	KR-20 Before Adj.	KR-20 After Adj.	SEM Before Adj.	SEM After Adj.
5-opt.	38	45	18.55	5.27	0.71	0.71	2.84	2.84
4-opt.	38	56	20.47	5.80	0.75	0.79	2.91	2.67
3-opt.	38	75	24.11	4.88	0.65	0.76	2.87	2.39
2-opt.	40	113	28.63	3.81	0.42	0.64	2.89	2.29